

Quantitative Correlation of Conformational Binding Enthalpy with Substrate Specificity of Serine Proteases

Birgit J. Waldner,[†] Julian E. Fuchs,^{†,‡} Roland G. Huber,^{†,§} Susanne von Grafenstein,[†] Michael Schauerperl,[†] Christian Kramer,[†] and Klaus R. Liedl^{*,†}

[†]Institute of General, Inorganic and Theoretical Chemistry, University of Innsbruck, Innrain 82, 6020 Innsbruck, Austria

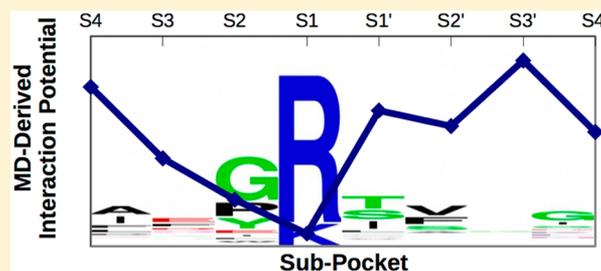
[‡]Centre for Molecular Informatics, Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom

[§]Bioinformatics Institute (BII), Agency of Science, Technology and Research (A* STAR), 30 Biopolis Street, Matrix#07-01, 138671 Singapore

Supporting Information

ABSTRACT: Members of the same protease family show different substrate specificity, even if they share identical folds, depending on the physiological processes they are part of. Here, we investigate the key factors for subpocket and global specificity of factor Xa, elastase, and granzyme B which despite all being serine proteases and sharing the chymotrypsin-fold show distinct substrate specificity profiles. We determined subpocket interaction potentials with GRID for static X-ray structures and an *in silico* generated ensemble of conformations. Subpocket interaction potentials determined for static X-ray structures turned out to be insufficient

to explain serine protease specificity for all subpockets. Therefore, we generated conformational ensembles using molecular dynamics simulations. We identified representative binding site conformations using distance-based hierarchical agglomerative clustering and determined subpocket interaction potentials for each representative conformation of the binding site. Considering the differences in subpocket interaction potentials for these representative conformations as well as their abundance allowed us to quantitatively explain subpocket specificity for the nonprime side for all three example proteases on a molecular level. The methods to identify key regions determining subpocket specificity introduced in this study are directly applicable to other serine proteases, and the results provide starting points for new strategies in rational drug design.



INTRODUCTION

Proteases are enzymes that catalyze the cleavage of peptide bonds and are important in numerous fundamental cellular processes. More than 550 proteases have been identified in the human genome, and about 100 more are predicted. Proteases also account for 1–5% of the genome of infectious organisms such as bacteria, parasites, and viruses, rendering them attractive drug targets.¹ Of all known proteolytic enzymes, more than one-third are serine proteases.²

While the mechanism of serine protease catalysis has been extensively studied,^{3,4} the key drivers of serine protease specificity have not yet been identified.⁵

To elucidate the molecular determinants of serine protease substrate recognition, we chose three serine proteases with chymotrypsin-fold adopting different biological functions and showing distinct specificity profiles. Substrate recognition occurs mainly in eight subpockets termed S4–S4' according to the convention of Schechter and Berger.⁶

Factor Xa (fXa) is an essential enzyme in the blood coagulation cascade where it cleaves prothrombin to thrombin and activates fVII.⁷ Its key role in the blood clotting process

makes it a target for anticoagulant drugs, and many of them are already in use.^{8,9}

According to the MEROPS cleavage site sequence logo,¹⁰ fXa prefers positively charged amino acids in the S1 pocket with a preference for Arg over Lys. In the S2 pocket fXa prefers small nonpolar amino acids such as Gly and Pro while the S3 pocket is rather unspecific. In the S4 pocket nonpolar amino acids such as Ile, Ala, and Phe are preferred. At the prime site in the S1' pocket mostly polar amino acids such as Ser and Thr are found. The S2'–S4' pockets all prefer nonpolar amino acids, with the S2' pocket showing a preference for Val according to MEROPS data.

Elastases are a group of proteases which cleave the important connective tissue protein elastin.¹¹ The here investigated porcine pancreatic elastase is structurally similar to human leukocyte elastase¹² and preferentially cleaves C-terminal amino acids with small alkyl side chains such as Ile, Val, and Ala.¹³ Because elastases can destroy connective tissue proteins and

Received: October 30, 2015

Revised: December 28, 2015

Published: December 28, 2015

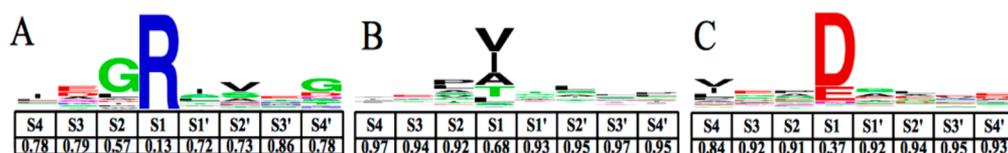


Figure 1. MEROPS cleavage site sequence logo and subpocket cleavage entropies of fXa (A), elastase (B), and granzyme B (C). The height of the single letter amino acid code indicates the preference for this amino acid in the corresponding subpocket. The cleavage entropy quantifies specificity in each subpocket S4–S4'. The heights of the single letter amino acid codes for elastase and granzyme B were upscaled for better visibility of the amino acid preferences.

may thus be very destructive if they are not regulated, they are controlled by either compartmentalization or naturally circulating plasma protease inhibitors.¹²

Granzyme B plays a key role in cytotoxic T lymphocyte mediated apoptosis¹⁴ and also shows antiviral and antitumor functions.¹⁵ Granzyme B is unique among mammalian serine proteases and strictly requires an aspartic acid in P1 position of substrates similar to caspases.¹⁶ Additionally granzyme B requires extended substrate interactions with preferences for Ile and Val at P4, Glu, Met or Gln at P3, broad preference at P2, an uncharged residue at P1', and Gly or Ala at P2'.¹⁷

Subpocket specificity can be quantified as cleavage entropy, a metric previously developed in our group.¹⁸ The cleavage entropy quantitatively reflects the cleavage site sequence logo from the MEROPS database (Figure 1 and Figures S1–S3 in the Supporting Information) and provides the basis for quantitative correlation of subpocket characteristics to subpocket specificity. Cleavage site sequence logos and cleavage entropies for all three example proteases are shown in Figure 1.¹⁹

In this study, starting from ligand-free X-ray structures of fXa, elastase, and granzyme B, we used a grid-based technique to probe enthalpic contributions to substrate specificity. In addition, we used molecular dynamics simulations to study the influence of conformational variability.

There are approaches to qualitatively explain fXa small molecule specificity^{20,21} and partially fXa substrate specificity.²² However, to our knowledge presently only the recent study of Raman and MacKerell²³ allows to quantitatively rationalize substrate specificity on a molecular level. In their recent work Raman and MacKerell quantitatively investigate thermodynamic contributions of the different molecular degrees of freedom for the binding of propane and methanol to fXa pockets. In contrast to our approach which is based on ligand-free protein states only, they use a thermodynamics-based endpoint method applied on a canonical ensemble of protein–ligand complexes as well as the corresponding free states.

Beyond proteases prediction of substrate promiscuity has been attempted before for G protein coupled receptors (GPCRs) by correlating various binding site descriptors to a set of descriptors for the corresponding ligands.²⁴ Approaches to explain hormone receptor specificity through simulation of specificity evolution have also been presented.²⁵

An *in silico* method for specificity profiling holds great promise to explore protease binding properties and in turn optimize success chances for molecular design targeting proteases that are key enzymes in many physiological pathways.^{26–28} The three presented examples share their chymotrypsin-fold with a wide range of other proteases and the methods introduced in this paper can thus directly be transferred to other protease systems with similar three-dimensional structure.

METHODS

Molecular Dynamics Simulations. The X-ray structures of fXa (PDB code 1C5M²⁹), elastase (PDB code 1QNJ³⁰), and granzyme B (PDB code 1FQ3³¹) were used as starting structures for molecular dynamics (MD) simulations as they are structures free of a ligand with a resolution <2 Å. Water molecules present in the X-ray structure within 4.5 Å from protein residues were kept to avoid removal of structural waters important for protein stability.³²

If present, the light chain was removed and protonation was carried out using the Protonate 3D tool in MOE.⁵ MD simulations in NTP conditions were performed for 1 μs using Amber 14³³ using the water model TIP3P³⁴ arranged in an octahedral solvation box with initial condition that the closest distance between any atom of the solute and the edge of the periodic box being 12 Å. Langevin dynamics with a collision frequency of 1 collision per picosecond were used for temperature control and constant pressure periodic boundary conditions with isotropic position scaling were applied.³⁵ Uniform neutralizing background plasma was used to neutralize the net charge of the periodic box. Long-range electrostatics were treated with particle mesh Ewald summation.³⁶ A cutoff of 8 Å was applied to nonbonding interactions. Equilibration was carried out using a previously developed protocol.³⁷ For the MD trajectory production run, a time step of 2 fs was used and snapshots were written out every 0.02 ns, resulting in 50 000 snapshots for 1 μs of MD trajectory. Bonds involving hydrogen bonds were constrained using the SHAKE algorithm.³⁸

Cpptraj³⁹ was used to analyze MD trajectories. Residue-wise root-mean-square fluctuations showed an overall agreement with experimental values calculated from the residue-wise temperature factors (obtained through averaging of the atom-wise temperature factors for each residue) from the X-ray structure.⁴⁰ Sampled active site conformations are compared to each other via two-dimensional RMSD plots (Figures S4–S6). Representative cluster conformations and cluster populations were determined through all atom RMSD based clustering of the binding site residues (see Tables S1–S3) of 25 000 equally spaced snapshots using the default hierarchical-agglomerative clustering algorithm⁴¹ implemented in cpptraj. The criterion applied to the selected number of clusters was that the occupancy for the least occupied cluster should not be below 1%. Granzyme B constitutes an exception as the least occupied cluster has only an occupancy of 0.4%. However, we wanted to have at least three representative cluster conformations for comparison purposes, independent of the cluster occupation. A structure at the center of each cluster in terms of all atom RMSD was selected to constitute the cluster's representative conformation.

Pocket Definition for Serine Proteases with Chymotrypsin-Fold. 32 structures of serine proteases with chymotrypsin-fold were downloaded from the PDB,⁴² and structures

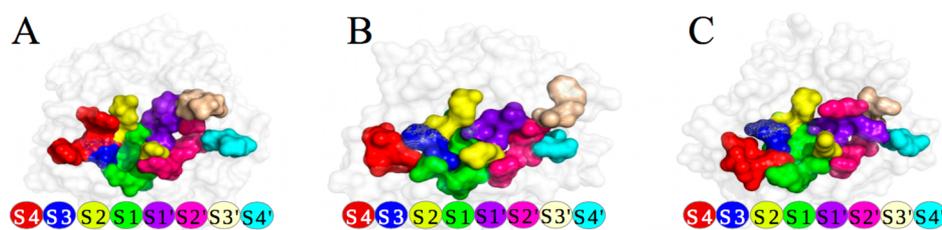


Figure 2. Pocket definitions for fXa (A), elastase (B), and granzyme B (C) based on the generic binding site definition for serine proteases with chymotrypsin-fold.

were aligned based on C- α atoms using PyMOL.⁴³ A list of X-ray structures used can be found in the Supporting Information (Text S1). A first pocket definition was created by selecting all enzyme residues within 4 Å proximity from the corresponding substrate residue. For each subpocket, the residues defining the subpockets were manually refined so that residues selected in other proteases with equivalent position were included in the definition and residues forming part of multiple subpockets excluded to minimize structural overlap between subpockets. Subpocket residues designated in the literature¹² were kept in the definitions even if they caused subpockets to overlap (Figure 2). It must be emphasized that the final pocket definition is a result of personal judgment of the authors. The residues of the pocket definitions for fXa, elastase, and granzyme B are given in Tables S1–S3.

Determination of Subpocket Interaction Potentials.

Local interaction potentials were determined with GRID⁴⁴ by placing a grid with 0.25 Å spacing over the whole binding site of each serine protease. The H₂O probe, the C3 probe, the O- probe, and the N3+ probe were used to calculate four different types of interaction maps. The H₂O probe is used to probe the subpockets for interactions with a hydrophilic substrate, while the C3 probe mimics interactions with a hydrophobic substrate. The N3+ probe and O- probe are used to test for electrostatic interactions.

Postprocessing of the GRID output files was carried out with an in-house C#-script. Grid points in proximity between 3.5 and 6 Å to at least two subpocket residues and an interaction potential <0 kcal/mol were selected in a first filtering step to include only grid points positioned within the subpocket and to avoid unfavorable positions due to van der Waals clashes. In a second filtering step only the 25% of points with the lowest interaction potentials were retained to represent the most favorable subpocket regions binding partners would select. Selected points were visualized using PyMOL.⁴³ Subpocket interaction potentials were calculated through averaging of the selected and filtered grid points.

Calculation of Weighted Mean of Subpocket Interaction Potentials and Correlation of Subpocket Interaction Potentials and Cleavage Entropy. Subpocket interaction potentials were determined for each representative cluster conformation. Cluster populations were used as weighting factors for the calculation of the weighted mean of subpocket interaction potentials.

Correlations between X-ray subpocket interaction potentials or weighted interaction potentials with the subpocket cleavage entropy were calculated as Pearson correlation coefficient r using the statistics package R.⁴⁵

RESULTS

The generated subpocket definition is derived from and applicable for various serine proteases. It allows to localize protein properties, like the here investigated interaction maps. The raw result from the described approach are filtered grid points localized at the individual subpockets for the starting structures as well as the representative conformations from the simulations, which are visualized in Figures S7–S125. It should be noted that definition of subpockets S4–S1 can be more reliably defined as there are more crystal structures available with substrates binding to S4–S1 than with substrates binding to all subpockets S4–S4'. Therefore, we focused the detailed description on S4–S1 only.

Interaction Maps for X-ray Structures. Factor Xa. The most favorable interactions in the S1 pocket are found with the N3+ probe, which interacts favorably with Asp-189, Gly-218, and Gly-226 at the bottom of the pocket and Asp-194 at the entrance of the pocket.

At the inside of the S2 pocket the O- probe shows the most favorable interactions with His-57, Gln-192, and Ser-214. The N3+ probe shows small regions of favorable interactions with Ser-214 and Trp-215.

In the S3 pocket favorable interactions are found between Ser-214 and the N3+ and H₂O probes. Some small spots with favorable interactions with the C3 probe are found, whereas there are no favorable areas found at the inside of the S3 pocket for the O- probe.

At the inside of the aromatic box in the S4 pocket of fXa only favorable interactions with the N3+ and the H₂O probe are found.

Elastase. In the S1 pocket mainly favorable interactions with the hydrophobic C3 probe are found close to Phe-228 and Cys-220. It is important to consider that the hydrophobic C3 probe systematically leads to smaller numerical values for this interaction than the interaction potentials with the O-, N3+, or H₂O probe because only van der Waals interactions are considered.

The hydrophobic C3 probe is preferred at the inside of the S2 pocket. Favorable interactions with the N3+ probe are found with Ser-214 and His-57 as well as with Phe-215. The O- probe shows some favorable interactions with His-57 and Gln-192.

Next to Val-216, the hydrophobic C3 probe is preferred in the S3 pocket. For the charged probes only a few points with considerable interaction potentials can be found.

No points with highly favorable interaction potentials are selected at the inside of the S4 pocket as the S4 assumes a closed conformation in the X-ray structure of elastase. The O- probe interacts favorably with Thr-175; some favorable interactions are also found between Thr-175 and the N3+ probe and H₂O probe.

Granzyme B. The S1 pocket in the granzyme B X-ray structure shows a clear preference for the O⁻ probe, which interacts favorably mainly with Ser-195, Asn-219, Arg-226. The N3⁺ probe shows some favorable interactions with Asp-194 at the entrance of the pocket. In comparison with the S1 pockets of the X-ray structures of elastase and fXa, the S1 pocket of granzyme B clearly prefers the O⁻ probe.

The O⁻ probe is preferred in the S2 pocket as well and interacts favorably with Lys-192, Ser-214, and Tyr-215. The S2 pocket of granzyme B shows a similar interaction profile as the S2 pocket of fXa with an even stronger preference for the O⁻ probe.

Also in the S3 pocket of granzyme B, the preference for the O⁻ probe is dominant. Small areas with favorable interactions with the H₂O and the C3 probe are also found. In comparison, fXa prefers the N3⁺ probe in the S3 pocket while elastase does not show any particular preferences.

In the X-ray structure, the S4 pocket assumes a rather closed conformation with some preferences for the O⁻ and the C3 probe close to Leu-171. The interactions found in the S4 pocket of granzyme B are not comparable to the interactions found in the aromatic box of fXa, but the interaction profile is similar to the one found for elastase.

Interaction Maps for Representative Conformations.
Factor Xa. Clustering of the MD trajectory for fXa results in six representative conformations with two more dominant clusters having occupancies of 43.7 and 20.8%, three less dominant clusters having occupancies of 15.3, 10.2, and 9.1%, and one minor cluster with an occupancy of 0.9%.

In all representative conformations the N3⁺ probe is clearly preferred in the S1 pocket. Major conformational changes between the different representative cluster conformations can be seen for the disulfide bridge between Cys-191 and Cys-220, Ala-190, and Tyr-228 which control opening and closing of the pocket.

The S2 pocket shows preferences for both the O⁻ and the N3⁺ probe at the inside of the pocket, depending on the conformations of Gln-192, Ser-214, and Trp-215. The rotamer of His-57 stays fixed in all six representative cluster conformations.

In the S3 pocket the rearrangement of Trp-215 in the representative conformations of clusters 2 and 3 in comparison to the representative conformation of cluster 1 allows for more favorable interactions with the O⁻ probe at the inside of the pocket. For the N3⁺ probe consistent favorable interactions with Ser-215 and the backbone of Trp-215 can be found in all representative cluster conformations.

The S4 pocket adopts a closed conformation in the representative conformation of cluster 1 with favorable interactions with the H₂O and N3⁺ probe in proximity to Tyr-85. The side chains of the residues of the aromatic box are shifted in other representative cluster conformations; however, only in cluster 4 they adopt positions causing an opening of the S4 pocket.

Elastase. Three representative cluster conformations were extracted from the MD trajectory of elastase. The clustering procedure revealed one predominant cluster with an occupancy of 75.6%, one less occupied with an occupancy of 19.4%, and a minor cluster with an occupancy of 5%.

The representative conformation of cluster 1 shows the S1 pocket in an open conformation with wide areas with favorable interactions with the hydrophobic C3 probe in proximity to Gln-190 and Phe-228. In the representative conformations of

clusters 2 and 3, the disulfide bridge between Cys-191 and Cys-220 is shifted in comparison to the representative conformation of cluster 1, causing a narrowing of the subpocket. Thr-213 is flipped in the representative conformations of clusters 2 and 3 which is reflected in different interaction maps with the charged O⁻ and N3⁺ probe. The largest difference between representative cluster conformations can be seen for Ser-217 which in the representative conformation of cluster 2 causes a closing of the binding pocket, while in the representative conformation of cluster 3 the S1 pocket is opened up with a different hydrogen bonding interaction profile due to the different position of Ser-217.

In the representative conformations of clusters 1 and 3, the side-chain of Gln-192 is oriented to point outward of the S2 pocket, which leads to less interactions with the O⁻ probe at the inside of the pocket. Also for Ser-214, the representative conformation of cluster 2 leads to an increase in favorable interactions with the O⁻ probe at the inside of the pocket. In the representative conformations of clusters 1 and 3 Ser-214 is oriented in a way that more favorable interactions are found at the edge of the pocket.

In the S3 pocket in the representative conformation of clusters 2 and 3 both Phe-215 and Val-216 are slightly shifted to adopt a conformation with less favorable interactions with the hydrophobic C3 probe. More favorable interactions with the N3⁺ and the O⁻ probe due to the increased backbone accessibility of Phe-215 for hydrogen bonding are found.

The S4 pocket adopts a closed conformation in all cluster representatives. Strong interactions are observed for the N3⁺ probe with Val-176 in the representative conformation of cluster 1 and with the O⁻ probe with Thr-175 in the representative conformation of cluster 3.

Granzyme B. As for elastase, three representative cluster conformations were extracted from the MD trajectory. Clustering revealed one predominant cluster with an occupancy of 89.1% and a less dominant cluster with an occupancy of 10.5%. The third cluster has only a very low occupancy of 0.4%.

In the representative conformation of cluster 1, the O⁻ probe is clearly preferred in the S1 pocket. In the representative conformation of cluster 2, Ser-190 and in the representative conformation of cluster 3 Arg-217 are shifted to cause a narrowing of the pocket and a decrease in the favorable area for the O⁻ probe.

In the representative conformation of cluster 1, the O⁻ probe and the H₂O probe are clearly preferred at the inside of the S2 pocket. In the representative conformation of cluster 2, Ser-214 and Tyr-215 are shifted and close the pocket. By contrast in the representative conformation of cluster 3, Ser-214 and the backbone of Tyr-215 are further rearranged so that the pocket adopts a more open conformation than in the representative conformation of cluster 1.

In all representative cluster conformations, the O⁻ probe and the H₂O probe are preferred in the S3 pocket. Only very few points of favorable interactions with the N3⁺ and the C3 probe can be found at the inspected energy levels.

In the representative conformation of cluster 1, the H₂O and the O⁻ probe are preferred in proximity to Arg-217. Only small areas with favorable interactions with the N3⁺ and the C3 probe are found at the inside of the S4 pocket. In the representative conformation of cluster 2 Arg-217 is shifted which opens the pocket and allows favorable interactions with the O⁻, C3, and H₂O probe at the inside of the pocket.

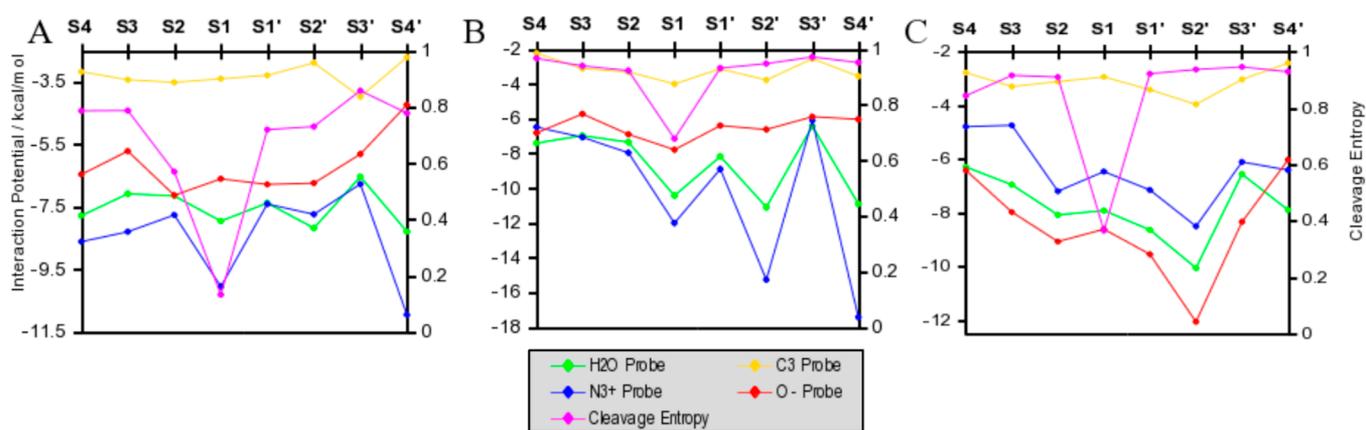


Figure 3. Interaction Potentials for X-ray structures of fXa (A), elastase (B), and granzyme B (C). It can be seen that results are more consistent for the nonprime site (subpockets S4–S1) than for the prime site (subpockets S1'–S4'). The strongest correlation between subpocket interaction potentials and cleavage entropy can be seen for fXa.

Quantitative Correlation between Interaction Potentials and Cleavage Entropy. Factor Xa. For a quantitative comparison with cleavage entropy an average value is derived from each interaction map (Figure 3A). The linear correlation between these interaction potentials calculated from the X-ray structure of fXa and the cleavage entropy is lower than 0.41 for all probes (Table 1). In the S4' pocket peaks for the interaction

Table 1. Correlation between Cleavage Entropy and Subpocket Interaction Potentials^a

probe	pockets	protease		
		FXa	elastase	granzyme B
N3+	S4–S4' X-ray/MD	0.41/0.84	0.19/0.27	–0.07/0.15
	S4–S1' X-ray/MD	0.76/0.83	0.94/0.84	0.11/0.35
	S4–S1 X-ray/MD	0.77/0.83	0.99/0.95	0.31/0.59
C3	S4–S4' X-ray/MD	0.06/0.51	0.61/0.58	–0.21/0.11
	S4–S1' X-ray/MD	0.37/0.53	0.83/0.90	–0.48/0.32
	S4–S1 X-ray/MD	0.31/0.49	0.84/0.90	–0.40/0.87
H ₂ O	S4–S4' X-ray/MD	0.29/0.79	0.40/0.60	–0.04/0.19
	S4–S1' X-ray/MD	0.57/0.83	0.95/0.84	0.10/0.40
	S4–S1 X-ray/MD	0.56/0.82	0.98/0.95	0.40/0.78
O-	S4–S4' X-ray/MD	0.37/0.46	0.78/0.84	–0.05/0.18
	S4–S1' X-ray/MD	0.33/0.54	0.77/0.99	0.02/0.20
	S4–S1 X-ray/MD	0.42/0.52	0.76/1.00	0.24/0.38

^aCorrelations are shown for X-ray structure and weighted average of subpocket interaction potentials using representative cluster structures obtained through MD simulations. The correlation coefficient r increases for each of the four GRID probes when using the weighted average of normalized interaction potentials of representative cluster conformations obtained through MD simulations and looking at all subpockets S4–S4' and except for elastase also when looking at subpockets S4–S1' and S4–S1.

potentials for the N3+ and the H₂O probe are detected. If only subpockets S4–S1' and S4–S1 are considered the correlation increases to $r = 0.76$ and $r = 0.77$ for the N3+ probe, also for the other probes a slight increase can be observed.

For the MD results the representative conformations are considered through weighting with the occupancies of the respective cluster. The N3+ probe almost perfectly follows the curve for the cleavage entropy (Figure 4A). The peak in the S4' disappears due to a better distribution of the areas with favorable interactions with the N3+ probe for the different

representative conformations. The highest correlation between subpocket interaction potentials and the cleavage entropy is $r = 0.84$. The worst correlation is found for the O- probe with $r = 0.46$. Considering only subpockets S4–S1' and S4–S1 only leads to a slight improvement for the H₂O, C3, and O- probe but does not improve correlation values for the N3+ probe.

Elastase. When comparing the interaction potentials calculated from the X-ray structure to the cleavage entropy (Figure 3B), it can be seen that the N3+ probe shows the lowest interaction potentials for subpockets S2–S4'. The interaction potentials for the N3+ and the H₂O probe both show spikes in the S2' and S4' pocket. Both subpockets show strong hydrogen bonding interactions with the N3+ and H₂O probe, but the number of grid points selected after the two filtering steps is almost a factor of 10 smaller than in other subpockets. The values for these two subpockets are thus not directly comparable to the results for the other subpockets as the number of points is not reflected in the average interaction potentials.

The correlation between subpocket interaction potentials is highest for the C3 and the O- probe when looking at all subpockets S4–S4'. If only subpockets S4–S1' or subpockets S4–S1 are considered, however, the correlation between the interaction potentials for the N3+ and H₂O probe increases to values $r > 0.94$. Also, the correlation between the interaction potential for the C3 probe increases, while the correlation between the O- probe and the cleavage entropy even shows a slight decrease if fewer subpockets are considered.

When looking at the weighted average of interaction potentials for representative cluster conformations, one sees that the peak in the S2' pocket disappears (Figure 4B). This is because in the representative cluster conformations the pocket adopts a more open conformation, and more grid points are selected. For the S4' pocket, however still there is only a small local very strong interaction and the peaks for the interaction potentials for the N3+ and the H₂O probe persist. For S4 and S3 pockets the H₂O probe now shows the strongest interaction potentials. The N3+ probe shows stronger interactions than the O- probe in the S4 if the weighted average of the interaction potential is used; however, there is not much difference between the O- and the N3+ probe.

Granzyme B. For granzyme B the interaction potentials calculated from the X-ray structure even show an unexpected inverse correlation between cleavage entropy and interaction

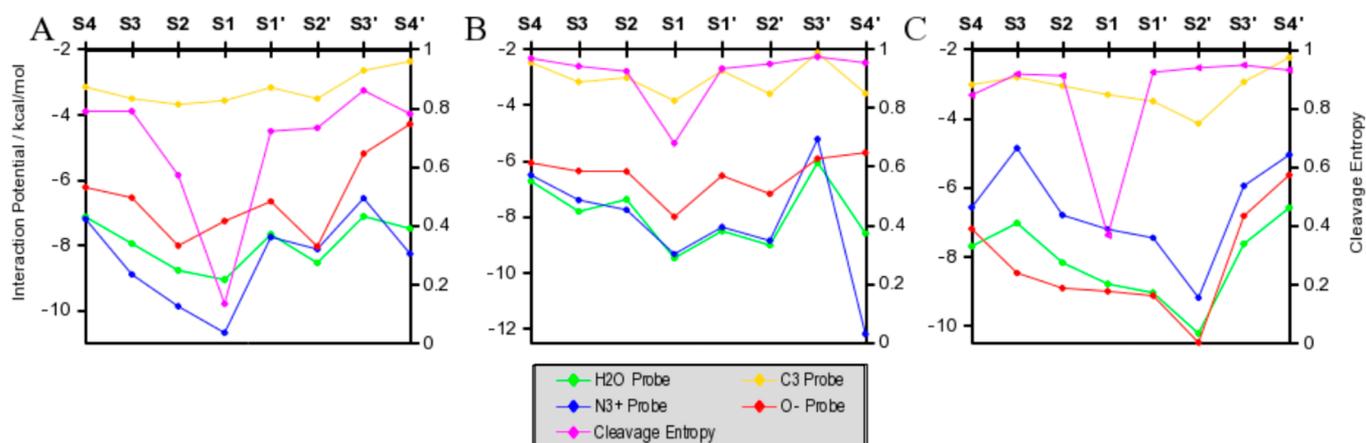


Figure 4. Weighted average of interaction potentials for representative cluster conformations of fXa (A), elastase (B), and granzyme B (C). It can be seen that correlation between subpocket interaction potentials and cleavage entropy is higher for the nonprime site (subpockets S4–S1) than for the prime site (subpockets S1'–S4'). The strongest correlation between subpocket interaction potentials and cleavage entropy can be seen for fXa.

potentials (Figure 3C). In the S2' and S3' only small local interaction areas are observed, meaning that a lower number of grid points is selected in the two filtering steps than in the other subpockets. For the S2' pocket this results in a peak of the interaction potential for the O- probe.

Usage of the weighted average of interaction potentials calculated from the three representative cluster conformations slightly improves the results (Figure 4C). Correlation coefficients between cleavage entropy and weighted average of interaction potentials now are $r > 0.11$ for all probes. Considering only subpockets S4–S1 increases the correlation between weighted average of interaction potentials and cleavage entropy considerably, looking only at subpockets S4–S1' doubles the correlation between weighted average of interaction potentials and cleavage entropy for the H₂O, C3, and N3+ probe.

DISCUSSION

Rationalizing Specificity through Consideration of Binding Site Conformational Variability. As shown by the correlation analysis presented here, the structural variability of proteases has to be considered to rationalize their substrate readout (Table 1). Subsequently, we will discuss the benefits of using a weighted average of subpocket interaction potentials determined from representative cluster conformations instead of subpocket interaction potentials obtained from the static X-ray structure.

Factor Xa. For fXa if only the interaction potentials obtained from the X-ray structure were considered, it could not be explained why small hydrophobic amino acids are preferred in the S4 pocket.

In the X-ray structure conformation the N3+ probe is preferred which can be explained by favorable cation– π interactions with the aromatic rings of the residues in the aromatic box.⁴⁶ Only when recognizing that the most abundant conformations show only a narrow S4 pocket with mostly no favorable interactions it can be explained that mainly Ala and Ile are found in fXa peptide substrates. The results are also in line with the finding that small cations are preferred in the S4 pocket due to cation– π interactions with small molecules.⁴⁷

The S3 pocket is rather small and influenced by the adjacent subpockets (S4, S2–S1). It shows no distinct interactions and

is thus accurately termed the hydrophobic pocket as only the C3 probe is favored at the inside of the pocket.

The preferences of the S2 pocket for hydrophobic residues and a minor preference for negatively charged residues can already be determined by looking at interaction potentials from the X-ray structure. However, only when investigating the different conformations it can be explained why the S2 pocket accepts mainly small hydrophobic amino acids such as Gly and Pro, but also occasionally Asp and Glu⁴⁸ which would only fit into the open conformations found in the representative conformations of clusters 3 and 6.

For the S1 pocket, specificity can already be explained from the X-ray structure as it shows such pronounced preferences for positively charged substrates. However, quantitative differentiation, especially from the O- probe is only possible when the other conformations are considered, because the affinity for positively charged substrates is enhanced in the most abundant conformations found in the MD simulation. The S1' pocket does not show major conformational changes in backbone or side chains during the MD simulation. Interactions with substrates in this subpocket are mainly due to hydrogen bonding to the protein backbone, which explains affinity for the N3+ probe, the H₂O probe, and the O- probe and the slight preference for Thr and Ser as amino acids in this subpocket as those amino acids both are able to form strong hydrogen bonds.

For the S2' pocket the results from the cluster analysis explain why hydrophobic residues are preferred but also why the subpocket accepts both positively charged and negatively charged amino acids.

While the representative conformation in cluster 2 allows for the accommodation of positively charged substrates, the representative conformation in cluster 3 prefers negatively charged ligands. For the S3' and S4' pockets a discussion in terms of chemistry is difficult, as these pockets cannot be accurately defined due to insufficient data both on substrate and X-ray structure side. The S3' pocket is rather unspecific which is reflected by both cleavage entropy and subpocket interaction potentials. In the applied definition of the S4' pocket both residues Thr-73 and Glu-74 allow for strong hydrogen bonding interactions, and the N3+ probe indicates salt bridges with the side chain of Glu-74, which leads to very strong localized

interaction potentials that are the cause for the peak in the S4' observed in Figures 3A and 4A.

Elastase. In the case of elastase which is a rather unspecific protease with substrate preferences only in the S1 subpocket, the usage of the weighted average of interaction potentials calculated from an ensemble of representative conformations leads to a decrease in the difference between the interaction potentials for the different probes. Considering that depending on the type of probe the interaction potentials are in a different range, Figure 3B shows that apart from the S1 pocket there are no distinct interactions which correctly reflects the substrate specificity as well as the cleavage entropy.

Granzyme B. For granzyme B, the weighted average of interaction potentials calculated from an ensemble of conformations allows to explain why Ile and Val are preferred in P4 position as there is virtually no difference between the N3+ and O- probe considering the possible ranges for interaction potentials of the different probes. They also more strongly depict the preference for negatively charged residues in P3 position as the O- probe is clearly preferred while the N3+ probe shows little interaction. However, results for both X-ray structure and ensemble of conformations disagree with the literature which infers substrate promiscuity at P2¹⁷ as the O- probe is preferred at the inside of the pocket in all cases.

For the S1, the preference for the O- probe is emphasized even more when the weighted average of interaction potentials calculated from an ensemble of conformations are used as all the representative cluster conformations adopt a more open conformation of the S1 than in the X-ray structure. Also, the preference for an uncharged residue at P1' can neither be explained by the X-ray interaction potentials nor the weighted average of interaction potentials.¹⁷ In S1' there are two charged residues allowing for the formation of salt bridges with charged probes (Arg-41 and Asp-194) and Ser-195 which also allows for the formation of strong hydrogen bonds.

Quantitative Correlation between Interaction Potentials and Cleavage Entropy. The GRID probe interaction potentials reflect the interaction potentials of equivalent moieties. Thus, charged GRID probes like the N3+ probe or the O- probe lead to stronger absolute interaction potentials than the H₂O probe or the C3 probe as electrostatic interactions are the strongest type of interactions observed between substrate and protein.⁴⁴

Cleavage entropy does not distinguish between different amino acid properties. Still, the interaction of positively charged substrates is dominating in the fXa cleavage profile of the S1 pocket. Thus, it is not surprising that for fXa the normalized interaction potential of the N3+ probe based on the X-ray structure shows the strongest correlation with the cleavage entropy ($r = 0.41$; see Figure 3A and Table 1).

Moreover, correlation is drastically improved when using the weighted average of interaction potentials based on the representative cluster conformations obtained through MD simulations. For the N3+ probe, usage of the weighted normalized interaction potentials leads to $r = 0.84$.

Also in the case of elastase, the usage of the weighted average of interaction potentials considerably improves correlation for all probes except the C3 probe when looking at all subpockets S4–S4'. For granzyme B correlation between weighted subpocket interaction potentials and cleavage entropy is still very low, but at least with weighted subpocket interaction potentials a positive correlation can be found between interaction potentials and cleavage entropy. The reasons for

the weakest results in the case of granzyme B might be that the subpockets of granzyme B are more different from most serine proteases with chymotrypsin-fold used for generation of the generic pocket definition. On the other side for granzyme B, the very prominent peak at S2' is a result from considering only few points with the selection algorithm which might result in a biased value as the number of points is not considered by the method.

The weights used for the calculation of weighted subpocket interaction potentials are not fitted, but rather derived directly from the cluster population. The strong correlation is remarkable as representative cluster conformations are obtained through MD simulations of ligand-free X-ray structures, whereas the cleavage entropy is calculated based on substrate data only.¹⁸ According to our findings, the presented method allows for prediction of the specificity profile from the intrinsic properties of the binding site. Our method distinguishes itself from existing cleavage site prediction approaches⁴⁹ as it is not based on machine learning algorithms^{50–52} and thus allows for direct physicochemical interpretation of the results.

Interestingly, correlations are improved in most cases when using the MD-derived interaction potentials independent of the type of probe. We attribute this effect to the importance of the underlying conformations. The interaction preferences of a subpocket are determined not only by the amino acids constituting the subpocket but also by the shape of the subpocket. The shape of the subpocket is considered implicitly through the selection procedure and averaging of selected grid points.

The weighted average of subpocket interaction potentials shown in Figure 4 thus depicts the ability of the pocket shape to interact with a substrate or probe. In case of a deeply buried pocket such as the S1 pocket, van der Waals interactions are possible from all sides which leads to more possibilities for specific interactions with a substrate and thus to higher specificity. If the pocket is more shallow or even convex, such as the S4 pocket in the sampled conformations, the possibilities for van der Waals interactions are limited, thus leading to less possibilities for specific interactions. This explains why in the case of fXa and elastase weighted interaction potentials for all probes correlate well with cleavage entropy. The effect is best described by looking at the results for the N3+ probe, but also the other probes depict the effect, as in each subpocket there are different regions with more or less favorable interaction potentials.

Recently, Duchene et al. applied a large-scale docking simulation to generate preference profiles for the two proteases matriptase and matriptase-2 and used the information to optimize peptidomimetic inhibitors.⁵³ In general, substrate-derived peptides and peptidomimetics are often the first step in the design of small molecule inhibitors.⁵⁴

In the case of fXa, the hydration thermodynamics have been studied by applying various methods.^{23,55,56} The study of Nguyen et al. applied scoring functions based on grid inhomogeneous solvation theory (GIST) to predict binding affinities. Results showed that a scoring function based on the energy of the displacement of water molecules performed as well as a scoring function based on both energy and water entropy. Thus, for fXa water entropy plays a minor role for binding affinity, but rather the energy of the displacement of water molecules has a major contribution to binding affinity.

Raman and MacKerell also have performed a spatial decomposition and thermodynamic quantification of driving

forces in fXa ligand specificity and observed both enthalpy–entropy compensation as well as reinforcement depending on the investigated binding pocket and ligand. However, they have found that direct protein–ligand interactions are significant for both polar and nonpolar binding and comparable to water reorganization energy.⁵⁷ These studies of water binding also support strong enthalpic interactions being the key to high substrate specificity in the case of fXa.

Our previous work mainly focused on entropic aspects, correlating specificity with the intrinsic flexibility of metalloproteases,⁵⁷ cysteine proteases,⁵⁸ and serine proteases.⁵⁹ The present work can be seen as complementary to our previous studies as we again confirm the importance of conformational variability for explaining substrate specificity. Additionally, by calculating interaction potentials with probe groups, we here include enthalpy as further aspect. Thus, a more complete picture of the thermodynamic contributions to substrate specificity of proteases could be obtained.

Our results show how important it is to use different conformations in virtual screening approaches and can be readily used in connection with flexible docking approaches.⁶⁰

CONCLUSIONS

Comparison of subpocket interaction potentials determined from the X-ray structure and weighted mean of subpocket interaction potentials calculated from representative cluster conformations showed a high dependence of subpocket interaction potentials on the investigated conformation. While subpocket interaction potentials determined from the X-ray structure showed little to no correlation to specificity quantified as cleavage entropy, subpocket interaction potentials determined as weighted mean from representative cluster conformations show high correlation to specificity in the case of elastase and fXa and minor correlation in the case of granzyme B. We therefore conclude that it is not sufficient to consider enthalpic contributions from a static structure, but one has to take into account dynamics in order to be able to accurately describe substrate recognition on a molecular level.

The methods developed for serine proteases are directly transferable to other proteolytic enzymes with similar structure and should also easily be transferable to other protease families. With the knowledge of the changes in possible substrate interactions depending on the conformation of the subpocket it may be possible to design inhibitors by performing chemical modifications that improve binding affinity.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpcc.5b10637.

A list of X-ray structures for generic pocket definition, pocket definition for fXa, elastase, and granzyme B, C- α RMSD values for X-ray structure and representative cluster conformations, cleavage site sequence logos for fXa, elastase, and granzyme B, all atom 2D-RMSD plots of binding site residues, figures with subpocket interaction potentials for X-ray structure and representative cluster conformations for all subpockets, comparison of experimental and theoretical RMSF values (PDF)

AUTHOR INFORMATION

Corresponding Author

*E-mail Klaus.Liedl@uibk.ac.at; Ph +43(512)507-57100; Fax +43(512)507-57199 (K.R.L.).

Present Address

C.K.: F. Hoffmann - La Roche AG, Grenzacherstrasse 124, 4070 Basel, Switzerland.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Birgit Waldner is thankful to the Austrian Academy of Sciences for being a recipient of the DOC grant. The authors also thank the Austrian Science Fund (FWF) for funding of project P 23051.

ABBREVIATIONS

fXa, factor Xa; GPCR, G protein coupled receptor; MD, molecular dynamics; GIST, grid inhomogeneous solvation theory; RMSD, root-mean-square deviation.

REFERENCES

- (1) Madala, S. K.; Pesce, J. T.; Ramalingam, T. R.; Wilson, M. S.; Minnicozzi, S.; Cheever, A. W.; Thompson, R. W.; Mentink-Kane, M. M.; Wynn, T. A. Matrix Metalloproteinase 12-Deficiency Augments Extracellular Matrix Degrading Metalloproteinases and Attenuates IL-13-Dependent Fibrosis. *J. Immunol.* **2010**, *184* (7), 3955–63.
- (2) Di Cera, E. Serine Proteases. *IUBMB Life* **2009**, *61* (5), 510–5.
- (3) Perona, J. J.; Craik, C. S. Structural Basis of Substrate Specificity in the Serine Proteases. *Protein Sci.* **1995**, *4* (3), 337–60.
- (4) Hedstrom, L. Serine Protease Mechanism and Specificity. *Chem. Rev.* **2002**, *102*, 4501–4524.
- (5) Di Cera, E. Engineering Protease Specificity Made Simple, but not Simpler. *Nat. Chem. Biol.* **2008**, *4* (5), 270–271.
- (6) Schechter, I.; Berger, A. Protease Subsite Nomenclature. *Biochem. Biophys. Res. Commun.* **1967**, *27*, 157–162.
- (7) Davie, E. W.; Fujikawa, K.; Kisiel, W. The Coagulation Cascade: Initiation, Maintenance, and Regulation. *Biochemistry* **1991**, *30* (43), 10363–10370.
- (8) Graff, J.; Harder, S. Anticoagulant Therapy with the Oral Direct Factor Xa Inhibitors Rivaroxaban, Apixaban and Edoxaban and the Thrombin Inhibitor Dabigatran Etxilate in Patients with Hepatic Impairment. *Clin. Pharmacokinet.* **2013**, *52* (4), 243–54.
- (9) Ewing, W. R.; Becker, M. R.; Manetta, V. E.; Davis, R. S.; Pauls, H. W.; Mason, H.; Choi-Sledeski, Y. M.; Green, D.; Cha, D.; Spada, A. P. Design and Structure-Activity Relationships of Potent and Selective Inhibitors of Blood Coagulation Factor Xa. *J. Med. Chem.* **1999**, *42*, 3557–3571.
- (10) Rawlings, N. D.; Barrett, A. J.; Bateman, A. MEROPS: The Database of Proteolytic Enzymes, their Substrates and Inhibitors. *Nucleic Acids Res.* **2012**, *40* (Database issue), D343–D350.
- (11) Bruch, M.; Bieth, J. G. Influence of Elastin on the Inhibition of Leucocyte Elastase by Alpha 1-Proteinase Inhibitor and Bronchial Inhibitor. Potent Inhibition of Elastin-Bound Elastase by Bronchial Inhibitor. *Biochem. J.* **1986**, *238* (1), 269–273.
- (12) Bode, W.; Meyer, E.; Powers, J. C. Human Leukocyte and Porcine Pancreatic Elastase: X-Ray Crystal Structures, Mechanism, Substrate Specificity, and Mechanism-Based Inhibitors. *Biochemistry* **1989**, *28* (5), 1951–1963.
- (13) Powers, J. C.; Gupton, B. F.; Harley, A. D.; Nishino, N.; Whitley, R. J. Specificity of Porcine Pancreatic Elastase, Human Leukocyte Elastase and Cathepsin G Inhibition with Peptide Chloromethyl Ketones. *Biochim. Biophys. Acta* **1977**, *485* (1), 156–166.
- (14) Lord, S. J.; Rajotte, R. V.; Korbitt, G. S.; Bleackley, R. C.; Granzyme, B. A Natural Born Killer. *Immunol. Rev.* **2003**, *193*, 31–38.

- (15) Trapani, J. A.; Sutton, V. R. Granzyme B: Pro-Apoptotic, Antiviral and Antitumor Functions. *Curr. Opin. Immunol.* **2003**, *15* (5), 533–43.
- (16) Fuchs, J. E.; von Grafenstein, S.; Huber, R. G.; Kramer, C.; Liedl, K. R. Substrate-Driven Mapping of the Degradome by Comparison of Sequence Logos. *PLoS Comput. Biol.* **2013**, *9* (11), e1003353.
- (17) Waugh, S. M.; Harris, J. L.; Fletterick, R.; Craik, C. S. The Structure of the Pro-Apoptotic Protease Granzyme B Reveals the Molecular Determinants of its Specificity. *Nat. Struct. Biol.* **2000**, *7* (9), 762–5.
- (18) Fuchs, J. E.; von Grafenstein, S.; Huber, R. G.; Margreiter, M. A.; Spitzer, G. M.; Wallnoefer, H. G.; Liedl, K. R. Cleavage Entropy as Quantitative Measure of Protease Specificity. *PLoS Comput. Biol.* **2013**, *9* (4), e1003007.
- (19) Crooks, G. E.; Hon, G.; Chandonia, J. M.; Brenner, S. E. WebLogo: A Sequence Logo Generator. *Genome Res.* **2004**, *14* (6), 1188–90.
- (20) Brandstetter, H.; Kuhne, A.; Bode, W.; Huber, R.; von der Saal, W.; Wirthensohn, K.; Engh, R. A. X-Ray Structure of Active Site-Inhibited Clotting Factor Xa. Implications for Drug Design and Substrate Recognition. *J. Biol. Chem.* **1996**, *271* (47), 29988–92.
- (21) Kamata, K.; Kawamoto, H.; Honma, T.; Iwama, T.; Kim, S.-H. Structural Basis for Chemical Inhibition of Human Blood Coagulation Factor Xa. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 6630–6635.
- (22) Rezaie, A. R. Determinants of Specificity of Factor Xa Interaction with its Physiological Inhibitors. *Mini-Rev. Med. Chem.* **2006**, *6* (8), 859–65.
- (23) Raman, E. P.; MacKerell, A. D. Spatial Analysis and Quantification of the Thermodynamic Driving Forces in Protein–Ligand Binding: Binding Site Variability. *J. Am. Chem. Soc.* **2015**, *137* (7), 2608–2621.
- (24) Levit, A.; Beuming, T.; Krilov, G.; Sherman, W.; Niv, M. Y. Predicting GPCR Promiscuity Using Binding Site Features. *J. Chem. Inf. Model.* **2014**, *54* (1), 184–194.
- (25) Eick, G. N.; Colucci, J. K.; Harms, M. J.; Ortlund, E. A.; Thornton, J. W. Evolution of Minimal Specificity and Promiscuity in Steroid Hormone Receptors. *PLoS Genet.* **2012**, *8* (11), e1003072.
- (26) Bachovchin, D. A.; Cravatt, B. F. The Pharmacological Landscape and Therapeutic Potential of Serine Hydrolases. *Nat. Rev. Drug Discovery* **2012**, *11* (1), 52–68.
- (27) Turk, B. Targeting Proteases: Successes, Failures and Future Prospects. *Nat. Rev. Drug Discovery* **2006**, *5* (9), 785–99.
- (28) Kawasaki, Y.; Freire, E. Finding a Better Path to Drug Selectivity. *Drug Discovery Today* **2011**, *16*, 985–990.
- (29) Katz, B. A.; Mackman, R.; Luong, C.; Radika, K.; Martelli, A.; Sprengler, P. A.; Wang, J.; Chan, H.; Wong, L. Structural Basis for Selectivity of a Small Molecule, S1-Binding, Submicromolar Inhibitor of Urokinase-Type Plasminogen Activator. *Chem. Biol.* **2000**, *7*, 299–312.
- (30) Würtele, M.; Hahn, M.; Hilpert, K.; Hoehne, W. Atomic Resolution Structure of Native Porcine Pancreatic Elastase at 1.1 Å. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2000**, *56*, 520–523.
- (31) Estébanez Perpiñá, E.; Bode, W.; Huber, R. *Crystal Structure of Human Granzyme B* **2003**.
- (32) Wallnoefer, H. G.; Fox, T.; Liedl, K. R. Challenges for Computer Simulations in Drug Design. In *Kinetics and Dynamics*; Springer: 2010; pp 431–463.
- (33) Case, D.; Babin, V.; Berryman, J.; Betz, R.; Cai, Q.; Cerutti, D.; Cheatham, T.; Darden, T.; Duke, R.; Gohlke, H.; et al. Amber 14, 2014.
- (34) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926.
- (35) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Oxford University Press: 1989.
- (36) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An $N \log(N)$ Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
- (37) Wallnoefer, H. G.; Liedl, K. R.; Fox, T. A Challenging System: Free Energy Prediction for Factor Xa. *J. Comput. Chem.* **2011**, *32*, 1743–1752.
- (38) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comput. Phys.* **1977**, *23* (3), 327–341.
- (39) Roe, D. R.; Cheatham, T. E., III PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9* (7), 3084–3095.
- (40) Bren, U.; Oostenbrink, C. Cytochrome P450 3A4 Inhibition by Ketoconazole: Tackling the Problem of Ligand Cooperativity Using Molecular Dynamics Simulations and Free-Energy Calculations. *J. Chem. Inf. Model.* **2012**, *52* (6), 1573–1582.
- (41) Shao, J.; Tanner, S. W.; Thompson, N.; Cheatham, T. E. Clustering Molecular Dynamics Trajectories: 1. Characterizing the Performance of Different Clustering Algorithms. *J. Chem. Theory Comput.* **2007**, *3*, 2312–2334.
- (42) Berman, H.; Henrick, K.; Nakamura, H. Announcing the Worldwide Protein Data Bank. *Nat. Struct. Biol.* **2003**, *10* (12), 980–980.
- (43) DeLano, W. *The PyMOL Molecular Graphics System*; DeLano Scientific LLC: Palo Alto, CA, 2008.
- (44) Goodford, P. J. A Computational Procedure for Determining Energetically Favorable Binding Sites on Biologically Important Macromolecules. *J. Med. Chem.* **1985**, *28* (7), 849–857.
- (45) Team, R. D. C. R. - A language and environment for statistical computing. Vienna, Austria, 2013.
- (46) Salonen, L. M.; Bucher, C.; Banner, D. W.; Haap, W.; Mary, J.-L.; Benz, J.; Kuster, O.; Seiler, P.; Schweizer, W. B.; Diederich, F. Cation- π Interactions at the Active Site of Factor Xa: Dramatic Enhancement upon Stepwise N-Alkylation of Ammonium Ions. *Angew. Chem., Int. Ed.* **2009**, *48* (4), 811–814.
- (47) Salonen, L. M.; Holland, M. C.; Kaib, P. S.; Haap, W.; Benz, J.; Mary, J. L.; Kuster, O.; Schweizer, W. B.; Banner, D. W.; Diederich, F. Molecular Recognition at the Active Site of Factor Xa: Cation-Pi Interactions, Stacking on Planar Peptide Surfaces, and Replacement of Structural Water. *Chem. - Eur. J.* **2012**, *18* (1), 213–22.
- (48) Stubbs, M. T.; Huber, R.; Bode, W. Crystal Structures of Factor Xa Specific Inhibitors in Complex with Trypsin: Structural Grounds for Inhibition of Factor Xa and Selectivity against Thrombin. *FEBS Lett.* **1995**, *375* (1–2), 103–7.
- (49) Yang, Z. R.; Thomson, R. Bio-Basis Function Neural Network for Prediction of Protease Cleavage Sites in Proteins. *IEEE Trans. Neural. Netw.* **2005**, *16* (1), 263–74.
- (50) Donald, M.; Spiegelhalter, D. J.; Taylor, C. C.; John, C. *Machine Learning, Neural and Statistical Classification*; Ellis Horwood: 1994; p 289.
- (51) Cai, Y. D.; Liu, X. J.; Xu, X. B.; Chou, K. C. Support Vector Machines for Predicting HIV Protease Cleavage Sites in Protein. *J. Comput. Chem.* **2002**, *23* (2), 267–74.
- (52) Chou, K.-C. Prediction of Human Immunodeficiency Virus Protease Cleavage Sites in Proteins. *Anal. Biochem.* **1996**, *233*, 1–14.
- (53) Duchêne, D.; Colombo, E.; Désilets, A.; Boudreault, P.-L.; Leduc, R.; Marsault, E.; Najmanovich, R. Analysis of Subpocket Selectivity and Identification of Potent Selective Inhibitors for Matriptase and Matriptase-2. *J. Med. Chem.* **2014**, *57* (23), 10198–10204.
- (54) Drag, M.; Salvesen, G. S. Emerging Principles in Protease-Based Drug Discovery. *Nat. Rev. Drug Discovery* **2010**, *9* (9), 690–701.
- (55) Abel, R.; Young, T.; Farid, R.; Berne, B. J.; Friesner, R. A. The Role of the Active Site Solvent in the Thermodynamics of Factor Xa-Ligand Binding. *J. Am. Chem. Soc.* **2008**, *130*, 2817–2831.
- (56) Nguyen, C. N.; Young, T. K.; Gilson, M. K. Grid Inhomogeneous Solvation Theory: Hydration Structure and Thermodynamics of the Miniature Receptor Cucurbit [7] Uril. *J. Chem. Phys.* **2012**, *137*, 044101.
- (57) Wallnoefer, H. G.; Lingott, T.; Gutierrez, J. M.; Merfort, I.; Liedl, K. R. Backbone Flexibility Controls the Activity and Specificity

of a Protein-Protein Interface: Specificity in Snake Venom Metalloproteases. *J. Am. Chem. Soc.* **2010**, *132* (30), 10330–7.

(58) Fuchs, J. E.; Grafenstein, S.; Huber, R. G.; Wallnoefer, H. G.; Liedl, K. R. Specificity of a Protein–Protein Interface: Local Dynamics Direct Substrate Recognition of Effector Caspases. *Proteins: Struct., Funct., Genet.* **2014**, *82*, 546–555.

(59) Fuchs, J. E.; Huber, R. G.; Waldner, B. J.; Kahler, U.; von Grafenstein, S.; Kramer, C.; Liedl, K. R. Dynamics Govern Specificity of a Protein-Protein Interface: Substrate Recognition by Thrombin. *PLoS One* **2015**, *10* (10), e0140713.

(60) Rao, M. S.; Olson, A. J. Modelling of Factor Xa-Inhibitor Complexes: A Computational Flexible Docking Approach. *Proteins: Struct., Funct., Genet.* **1999**, *34* (2), 173–83.